



## Miranda

Revue pluridisciplinaire du monde anglophone /  
Multidisciplinary peer-reviewed journal on the English-speaking world

19 | 2019

Rethinking Laughter in Contemporary Anglophone Theatre

---

# Humanités numériques et études anglophones : Comprendre et explorer

Géraldine Castel

---



### Édition électronique

URL : <http://journals.openedition.org/miranda/20477>

DOI : 10.4000/miranda.20477

ISSN : 2108-6559

### Éditeur

Université Toulouse - Jean Jaurès

### Édition imprimée

Date de publication : 7 octobre 2019

### Référence électronique

Géraldine Castel, « Humanités numériques et études anglophones : Comprendre et explorer », *Miranda* [En ligne], 19 | 2019, mis en ligne le 08 octobre 2019, consulté le 16 février 2021. URL : <http://journals.openedition.org/miranda/20477> ; DOI : <https://doi.org/10.4000/miranda.20477>

---

Ce document a été généré automatiquement le 16 février 2021.



Miranda is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

---

# Humanités numériques et études anglophones : Comprendre et explorer

Géraldine Castel

---

*Les références des logiciels et projets évoqués sont récapitulées en fin d'article.*

- 1 À l'heure où les « humanités numériques » gagnent en visibilité dans le paysage de la recherche française et internationale, où des manifestes sont rédigés, des pôles et des chaires créés, des équipes rassemblées, des projets financés, faire ses premiers pas sur un terrain dont la définition même est objet de controverse peut s'avérer délicat.
- 2 Parfois qualifiée d'oxymore (Greetham, 532, 2012), l'association des deux termes, *digital humanities* en anglais puis « humanités numériques » en français, de même que sa signification font encore l'objet de nombreuses polémiques. Discipline ? Post-discipline ? (Welger-Barboza, 2012) Démarche ? Méthodologie ? Appareillage technique ? Zone de convergence ? Le débat n'est pas tranché même si les chercheurs appartenant à cette nébuleuse tentent régulièrement d'en délimiter les frontières (Nyhan, Terras & Vanhoutte, 2014). Pour les francophones, s'ajoute la question de la pertinence de la traduction littérale généralement adoptée, comme le souligne par exemple Guichard (Guichard, 13, 2014). *Humanities* en anglais et « humanités » en français ne renvoient notamment pas à des champs disciplinaires identiques.
- 3 Enfin, la clarté terminologique de l'expression n'est pas facilitée par l'approche délibérément ouverte associée à ce champ d'études, celle du « chapiteau » (*Big Tent*) choisie notamment comme thématique des rencontres annuelles des humanités numériques organisées en 2011 à l'université Stanford (DH2011). Si cette ouverture est en effet source de flexibilité et de richesse, elle n'est pas sans risque comme le souligne Melissa Terras (Terras, 2012) :

The field can only continue to expand, as more and more people engage with the technology that allows them to undertake academic tasks, and who are we to ring fence the academic field that lets people discuss this and learn more? [...] But if

everyone is a Digital Humanist, then no-one is really a Digital Humanist. The field does not exist if it is all pervasive, too widely spread, or ill defined.

- 4 De cette multiplicité d'interprétations et de définitions en évolution constante est ici retenue celle proposée par Marin Dacos et Pierre Mounier dans leur rapport réalisé à la demande de l'Institut Français et publié en 2014 (Dacos et Mounier, 15, 2014) :

Au plus haut niveau de généralité, on pourrait dire que les humanités numériques désignent un dialogue interdisciplinaire sur la dimension numérique des recherches en sciences humaines et sociales, au niveau des outils, des méthodes, des objets d'études et des modes de communication.

- 5 Un tel dialogue a bien sûr connu des précédents comme le rappelle Aurélien Berra (Berra, 2011) qui évoque notamment le travail de Roberto Busa, jésuite italien qui dès les années 40 entreprit d'automatiser l'indexation d'items issus des œuvres de Thomas d'Aquin, ou encore la création en 1963 du *Centre for Literary and Linguistic Computing* à Cambridge. Hélène Bourdeloie précise néanmoins à propos des humanités numériques contemporaines (Bourdeloie, 15, 2014) :

La véritable nouveauté réside dans le fait que le numérique n'est plus seulement un outil au service de la recherche mais aussi un objet de recherche à part entière. Il est à la fois instrument, méthode, terrain et objet de recherche.

- 6 Au-delà des effets de mode, ce territoire complexe et inégalement défriché peut présenter un potentiel réel pour les chercheurs souhaitant s'y aventurer. Potentiel financier puisque les appels à projets intégrant cette thématique se multiplient. Potentiel professionnel puisque l'intitulé apparaît de manière croissante dans les profils de postes universitaires mis au recrutement. Mais surtout potentiel scientifique, et ce, quelles que soient les disciplines concernées.

- 7 Dans les études anglophones françaises, des chercheurs ont commencé à investir ce terrain à l'image de Grégoire Lacaze d'Aix-Marseille Université qui a donné en 2018 une présentation intitulée « Du discours rapporté dans les Tweets : reconfiguration des pratiques citationnelles du discours circulant » ou encore de l'équipe de l'université Paris 8 ayant piloté le programme de recherche « Le sujet digital »<sup>1</sup>, dans le cadre duquel a été organisé le colloque « Posthumain et subjectivités numériques »<sup>2</sup> en 2016. Claire Larssonneur, membre de cette équipe et spécialiste de traduction et de littérature, a par ailleurs coordonné en 2018 un numéro de la revue *Angles* de la SAES intitulé « Digital Subjectivities ». Pour autant, comme en témoigne par exemple le programme du 15e colloque de l'Institut des Amériques consacré aux humanités numériques qui s'est tenu à la Rochelle en 2017, les anglicistes français s'intéressant à ces thématiques sont encore relativement peu nombreux.

- 8 Le Manifeste des humanités numériques rédigé de manière collaborative suite à la tenue de séminaires dirigés par Todd Presner (UCLA) et Jeffrey Schnapp (Stanford) définissait les objectifs du mouvement de la manière suivante (*Digital Humanities Manifesto 2.0*, 2009) :

The revolution is not about transforming literary scholars into engineers or programmers. Rather, it is about:

- expanding the compass and quality of knowledge in the human sciences
- expanding the reach and impact of knowledge in the Humanities disciplines
- direct engagement in design and development processes that give rise to richer, multidirectional models, genres, iterations of scholarly communication and practice.

- 9 Mais si ces ambitions ne sont pas utopistes (Cf Greenspan, 2016), quelles formes peuvent-elles donc prendre en pratique ? Via quelles applications concrètes ? Dans un champ aussi hétérogène et encore en construction, il n'y a pas de réponse unique à ces questions mais une multitude de voies expérimentales ouvertes par des acteurs divers dessinant une mosaïque complexe. L'article ci-dessous vise à en proposer un aperçu.
- 10 Pour ceux qui ont décidé de se lancer, de multiples ouvrages détaillés peuvent accompagner les chercheurs désireux de pénétrer ce territoire (Crompton & Lane, 2016 / Gardiner & Musto, 2015), utilement complétés par des réflexions à l'échelle des SHS françaises (Barats, 2013 / Clavert *et al.*, 2017). Néanmoins, pour ceux qui ne perçoivent pas l'intérêt scientifique d'ajouter cette corde à leur arc, il semble utile de proposer ici une introduction synthétique des apports potentiels du numérique aux différentes étapes du travail de recherche, illustrée d'exemples tirés des disciplines des études anglophones françaises afin d'informer la réflexion des chercheurs du domaine.

## Cycle de recherche et numérique : Les sources

- 11 En premier lieu, l'adoption des technologies de l'information et de la communication par différents protagonistes institutionnels ou issus de secteurs comme ceux de la documentation, de la presse et de l'édition, a permis de grandement faciliter l'accès aux sources. La bibliothèque nationale d'Écosse propose par exemple une archive appelée *The Word on the Street* comprenant environ 1800 *broadsides*, des documents imprimés d'une page destinés à être affichés publiquement, datant de 1650 à 1910, au contenu très varié. Un moteur de recherche et un classement par thématique permettent d'identifier les sources pertinentes qu'il est ensuite possible de télécharger au format PDF.
- 12 À une échelle plus large, celle de l'Union européenne, a été créée l'infrastructure de recherche CLARIN (*Common Language Resources and Technology Infrastructure*) ayant pour vocation de rassembler en une seule plateforme une multitude de jeux de données destinés aux chercheurs en sciences humaines et sociales. Des sources utiles à des anglicistes y sont proposées, d'une archive de tweets en gallois contenant sept millions d'items (*The Corpus of Welsh Language Tweets*) à un corpus de journaux anglais des 17<sup>e</sup> et 18<sup>e</sup> siècles ciblant des linguistes travaillant sur l'histoire de la langue anglaise (*Zurich English Newspaper Corpus*) en passant par la totalité des romans de Charles Dickens (CLiC).
- 13 Au-delà des institutions patrimoniales, gouvernementales etc., et parfois en partenariat avec ces institutions se développent des programmes destinés à faciliter l'accès à des sources hétérogènes, comme en témoignent notamment des initiatives pionnières telles que *The Valley of the Shadow* ou l'archive Rossetti. La première, lancée à la fin des années 90, met à disposition de nombreux documents numérisés tels que lettres, journaux, cartes, discours et rapports statistiques ou de recensement relatifs à deux communautés engagées dans la guerre de Sécession américaine. La seconde, créée en 2000, offre un panel très large de ressources portant sur l'œuvre de Gabriel Dante Rossetti, artiste britannique polyvalent du 19<sup>e</sup> siècle : manuscrits, correspondance, dessins et peintures mais aussi tout un écosystème de documents relatifs à son travail tels que les textes d'origine de ses traductions, les illustrations accompagnant ses poèmes ou encore les articles publiés sur ses travaux. Des outils sont par ailleurs

proposés pour permettre l'exploitation de ces ressources tels que moteur de recherche, application permettant d'élaborer des collections et des expositions en ligne ainsi qu'une plateforme de jeu numérique, aujourd'hui désactivée mais initialement destinée à conduire des exercices d'interprétation critique de documents de façon collaborative.

- 14 Les archives de ce type, appuyées sur des bases de données qui gagnent progressivement en richesse et en hétérogénéité, sont de plus en plus nombreuses et réalisées à l'initiative de différents acteurs, privés ou publics. Elles présentent une valeur scientifique évidente pour la communauté de la recherche dans une variété croissante de domaines, qu'il s'agisse de littérature, de civilisation, mais aussi de didactique, de linguistique ou encore de phonétique. Ainsi, le *Newcastle Electronic Corpus of Tyneside English* par exemple met à disposition des documents relatifs à ce dialecte tels que versions audio numérisées, transcriptions orthographiques ou phonétiques.
- 15 Certaines archives mènent également une mission de conservation de ressources menacées comme par exemple le réseau DELAMAN (*Digital Endangered Languages and Musics Archive Network*) et son projet *Lost & Found* destiné à identifier des sources linguistiques risquant la disparition sans un effort de numérisation concerté et à compléter ces sources par des métadonnées pertinentes.
- 16 À l'origine, ces archives proposaient essentiellement la numérisation ou l'encodage de documents non-numériques. Mais l'activité humaine migrant en ligne dans de nombreux domaines, elles tendent désormais à inclure des documents dits natifs, générés directement en ligne, sans équivalent hors-ligne et dont la préservation devient elle aussi essentielle pour lutter contre le phénomène de détérioration du web (*Web Decay*) souligné par exemple par Tom Chatfield (Chatfield, 2014).
- 17 C'est l'objectif notamment de l'archive de l'internet des droits de l'homme de l'université de Columbia (*Human Rights' Web Archive*) qui sauvegarde depuis 2008 les pages de centaines de sites liés à cette thématique (ONG, Tribunaux, institutions, individus...) dans une dizaine de langues ou encore de la machine à remonter le temps (*Wayback Machine*) de l'association à but non lucratif *The Internet Archive*, basée aux États-Unis et qui effectue des copies de millions de pages web à l'échelle mondiale pour construire une « bibliothèque de l'internet » à transmettre aux générations futures. La fréquence de collecte de ces captures est inégale de même que la qualité des pages mises à disposition mais cet outil peut néanmoins s'avérer précieux pour accéder à des contenus disparus du web contemporain.<sup>3</sup>
- 18 Ces documents natifs sont souvent très riches sur le plan quantitatif et qualitatif. Ainsi, l'archive numérique du 11 septembre (*September 11 Digital Archive*) recense et met à disposition environ 150.000 témoignages de toute personne souhaitant s'exprimer sur cet événement sous différentes formes : emails, photos, vidéos ou encore enregistrements audio. Sur le même principe est proposée la *Canterbury Earthquake Digital Archive* (CEISMIC) relative aux tremblements de terre de 2010 et 2011 en Nouvelle- Zélande. Ce principe de production participative est également utilisé dans l'archive *Scots Words and Places* de l'université de Glasgow offrant un glossaire de mots écossais renvoyant à des lieux et proposant leur forme ancienne, leur signification et des exemples d'utilisation.
- 19 Ces données numériques natives offrent ainsi par exemple la possibilité d'étudier la perception de différentes œuvres littéraires ou télévisées auprès de leurs publics, comme le propose une étudiante de l'université Northeastern de Boston (*Operation Critique: The Changing Face of the Literary Conversation*) à partir des publications de

lecteurs ou téléspectateurs sur la plateforme Tumblr. Les réseaux sociaux font par ailleurs l'objet d'études visant notamment à définir des modèles pour une exploitation pédagogique et didactique de ces supports (Burnett & Merchant, 2011).

- 20 Ces archives numériques offrent donc une quantité importante de documents sous de multiples formes comme en témoignent par exemple les ressources proposées par la *British Library* : images de contenus tombés dans le domaine public mises à disposition sur Flickr, enregistrements audio sur la page *Accents and Dialects* à visée phonétique mais aussi série d'entretiens avec des auteurs bilingues intitulée *Between Two Worlds, Poetry and Translation* à vocation littéraire ou encore le projet *Observing the 1980s* rassemblant des entretiens réalisés auprès de britanniques aux profils très divers sur les thématiques clés de la décennie : grève des mineurs, montée de la contamination au VIH, désindustrialisation du pays... La vidéo est également très présente grâce à des plateformes telles que celle du centre Kennedy (*Kennedy center Millennium Stage Performance Archive*). Ces archives numériques facilitent ainsi l'accès à des sources dont le caractère hétérogène enrichit indéniablement l'exploitation scientifique, à l'image par exemple des projets réalisés à partir de l'archive *Berkeley Revolution* mise à disposition par l'université de Californie. L'un d'entre eux suit ainsi le parcours d'une étudiante ayant fui sa famille pour rejoindre le campus à partir d'une lettre écrite à sa mère mais jamais postée saisie par la police et publiée en partie dans la *Berkeley Gazette* en 1970, puis de son journal intime et enfin, d'un long commentaire rédigé par la même femme quarante ans plus tard sur un blog.
- 21 Enfin, au-delà des ressources mises à disposition par différents fournisseurs, il devient également possible à chacun de se constituer son propre jeu de données numériques via des outils comme l'API (Interface applicative de programmation permettant l'échange de données) de Twitter ou encore l'extension *Save Webpage as Word Document* pour le navigateur Chrome.
- 22 La consultation de sources diverses a donc été facilitée par l'apparition du numérique mais leurs conditions d'accès et de réutilisation sont très variables. Le *Guardian* britannique a par exemple fait le choix d'ouvrir librement et gratuitement ses archives à des fins non commerciales. Les ressources comme la correspondance d'Abraham Lincoln ou les quelques sept mille affiches publicitaires de l'université Duke tombées dans le domaine public sont également disponibles sans frais. Pour autant, de nombreuses sources numériques ont une accessibilité restreinte, du fait de leur monétisation, de contraintes juridiques relatives aux droits d'auteur, au droit à la vie privée pour les données issues des réseaux sociaux par exemple, ou enfin de problématiques techniques liées à des questions de collecte, de stockage et de traitement de données souvent volumineuses, hétérogènes et interconnectées. Au-delà des questions d'accès, l'exploitation scientifique et le partage de données numériques peuvent également faire l'objet d'une réglementation souvent délicate à appréhender, en évolution constante et divergente d'un pays à un autre. Des initiatives ont cependant vu le jour pour accompagner les chercheurs sur ce volet à l'image par exemple du carnet *Éthique et droit* ouvert depuis 2011 sur la plateforme Hypothèses par un collectif de chercheurs, de professionnels de la documentation électronique et de juristes sous l'impulsion de Véronique Ginouvès et d'Isabelle Gras de l'université d'Aix-Marseille.

## Cycle de recherche et numérique : Analyse et exploitation

- 23 Une fois ces sources à disposition, ces objets d'étude numériques peuvent être exploités par le biais de méthodes traditionnelles mais les outils développés au cours des dernières décennies peuvent également intervenir à chaque étape du travail de recherche, de la réalisation d'un état de l'art à la diffusion des résultats d'un projet.<sup>4</sup>
- 24 En ce qui concerne ce premier travail, il est désormais courant d'utiliser *Google Scholar* lors de la phase d'identification de sources secondaires mais ce n'est pas le seul outil dans ce domaine. En effet, il existe aussi des moteurs tels que *CORE (Connecting REpositories)*, élaboré par le *Knowledge Media Institute* qui répertorie exclusivement des articles ou ouvrages disponibles en libre accès. En mai 2018, il recensait 131 millions de métadonnées et proposait au téléchargement 11 millions de textes complets (Notay, 2018). Ce moteur se distingue également de son plus célèbre concurrent par la volonté de ses concepteurs de non seulement identifier des articles mais également permettre leur collecte automatique pour ensuite pratiquer de l'exploration de données (Data Mining) sur le corpus ainsi créé. *BASE (Bielefeld Academic Search Engine)* quant à lui propose une fonction permettant de connaître pour chaque document répertorié ses conditions d'accès et de réutilisation.
- 25 Une fois les documents pertinents à la réalisation d'un état de l'art identifiés, il convient d'en dresser un inventaire avant de les consulter. C'est à ce stade que des outils tels que *Zotero* ou *Mendeley* peuvent offrir une aide précieuse. Le principe de fonctionnement est le suivant : lorsque l'utilisateur se rend sur un site compatible (Les moteurs de recherche précédemment évoqués, HAL, Amazon...) pour effectuer une recherche, *Zotero* reconnaît automatiquement les informations bibliographiques présentes sur la page et propose de les télécharger, voire d'y ajouter le PDF de la ressource choisie s'il est disponible. Ceci est ensuite stocké dans une bibliothèque numérique qu'il est alors possible de classer selon différents critères et thématiques. Les PDF peuvent être annotés directement dans le logiciel et des bibliographies automatiquement générées à partir du contenu de la bibliothèque selon le format choisi (MLA...). Il est enfin possible de partager le contenu de sa bibliothèque ou de le modifier de manière collaborative. Un tutoriel vidéo très complet réalisé par l'Urfist (Unité Régionale de Formation à l'Information Scientifique et Technique) est disponible sur Youtube. La qualité du résultat produit est conditionnée par la fiabilité des métadonnées disponibles en ligne sur les sites utilisés mais même si une relecture reste évidemment indispensable, le gain de temps est généralement considérable.
- 26 À l'étape suivante, lorsque le travail sur les données à proprement parler s'amorce, d'autres types de logiciels sont employés pour nettoyer si nécessaire puis organiser ces données afin de rendre possible l'analyse ultérieure, de discours ou de réseaux par exemple. *OpenRefine* en fait partie. Cet outil fonctionne de manière optimale avec des données sous forme tabulaire bien qu'il puisse prendre en charge d'autres formats. Il est particulièrement précieux lorsqu'un fichier comporte des erreurs de saisie, des coquilles, des cellules vides, des répétitions etc. sur un nombre important de données, ce qui rend la correction manuelle extrêmement chronophage. *OpenRefine* possède notamment une fonction clustering/regroupement qui permet de faciliter le travail de normalisation d'un document. En effet, lorsqu'un même fichier est le fruit d'une saisie collective, il arrive par exemple que des dates soient saisies dans des formats différents



(9.11.2011 / September 11, 2011...). S'il est demandé ensuite à un logiciel de classer ces dates dans l'ordre, il ne sera en mesure de le faire que si toutes ces dates apparaissent dans un format identique, d'où l'intérêt de ce type de manipulation. OpenRefine suggère ainsi des regroupements sur des données qu'il estime similaires et donne la possibilité d'unifier leur format en quelques clics.

- 27 D'autre part, OpenRefine permet également d'appliquer des filtres (appelées « facettes ») sur un jeu de données afin d'isoler des éléments recherchés. Sur un fichier comportant les métadonnées relatives à tous les romans du 19<sup>e</sup> siècle, il sera ainsi par exemple capable d'identifier rapidement tous ceux de plus de 100 pages publiés en Australie par une maison d'édition précise de 1850 à 1900, comprenant le mot « passion » et de créer un sous-jeu selon ces critères. Fait appréciable, OpenRefine conserve un historique complet des modifications apportées à un fichier qu'il est possible d'exporter pour référence ultérieure ou pour application à un autre document.
- 28 Ce type de requêtes peut également être réalisé via des systèmes de gestion de données (DBMS) qui permettent d'organiser, de classer et de stocker des données lorsque leur volume et leur complexité testent les limites d'un travail plus artisanal. Pour formuler ces requêtes (Ex : Les phrases exclamatives contenant le modal CAN énoncées par un locuteur irlandais entre 2005 et 2012), la plupart de ces systèmes fonctionnent grâce à un langage spécifique appelé SQL (*Structured Query Language*) mais il existe néanmoins des interfaces utilisateurs dotées d'un assistant qui ne nécessitent pas de compétences en programmation. C'est le cas par exemple de LibreOffice Base. L'utilisation qu'en fait Juliet Boyd est représentative des fonctionnalités d'un tel outil. En effet, selon un phénomène que les chercheurs en littérature notamment reconnaîtront, cet auteur d'une série de huit romans pour adolescents s'est progressivement retrouvée à devoir travailler à partir d'un nombre croissant de notes prises sur des supports divers qu'il devenait de plus en plus difficile de gérer manuellement. Elle a donc utilisé Base pour centraliser les informations relatives à chaque chapitre en recensant les personnages y apparaissant, l'intrigue, les lieux etc.. et contenant des liens vers des tableaux apportant davantage d'informations sur ces catégories selon le principe de la base de données relationnelle. Cet outil lui offre ainsi la possibilité de réconcilier vision synthétique des données et contenus détaillés.
- 29 Faire des recherches sur un corpus peut également être abordé via l'encodage. Il s'agit alors de poser des balises dans un texte pour identifier des éléments récurrents de contenu, de structure... personnalisables et permettre à terme une analyse automatisée. La plateforme *Curious Travellers* de l'université de Glasgow met ainsi à disposition des lettres et récits de voyage du 18<sup>e</sup> siècle essentiellement issus de la correspondance du naturaliste britannique Thomas Pennant. Dans ces textes, les lieux, les noms de personnes, les références à des livres et à des œuvres d'art sont encodées. Grâce à cet encodage, Ffion Mair Jones a ainsi pu étudier l'itinéraire des différents voyages de Pennant mais aussi sa perception de figures historiques telles que l'impératrice Catherine II de Russie.
- 30 Ce type d'encodage est généralement réalisé grâce au langage XML ou plus récemment, en Json, via un traitement de texte ou un logiciel spécialisé comme OXygen. Pour s'initier à ce type d'exercice, il existe des modèles pour différents types de contenus (Théâtre, linguistique...) sur la plateforme Roma mise à disposition par la TEI (*Text Encoding Initiative*). Pour faciliter ce type d'exercice, la TEI est un consortium international qui tente d'uniformiser les pratiques autour d'un référentiel de normes



communes. Si l'encodage peut être un exercice extrêmement chronophage, de nombreuses sources déjà encodées sont disponibles pour utilisation immédiate telles que la collection de textes néo-zélandais proposée par l'université Victoria de Wellington ou la collection de dictionnaires bilingues Freedict. Dans le cas de sources pré-encodées comme celles-ci, les choix opérés par l'équipe responsable de l'encodage ne coïncident pas nécessairement avec la problématique du chercheur qui souhaite les utiliser a posteriori. Elles peuvent néanmoins offrir un point d'entrée utile dans un jeu de données ou encore permettre de se familiariser avec les différentes manipulations rendues possibles par l'encodage.

- 31 Les logiciels d'annotations (*Tagging*) recourent en partie ce type de fonctionnalités. Il n'est plus question de poser des balises XML mais des étiquettes via un logiciel spécialisé. L'objectif est ici de permettre de résoudre la problématique rencontrée par de nombreux chercheurs. En effet, prendre des notes directement sur un document permet de conserver le texte sous les yeux mais devient vite illisible lorsque ces notes se multiplient. À l'inverse, noter des commentaires à part permet de conserver la lisibilité du document d'origine mais nécessite de garder systématiquement les deux en parallèle pour ne pas perdre le lien au texte, ce qui peut poser problème lorsque le nombre de textes travaillés devient important. Avec un logiciel dédié, les annotations apparaissent en parallèle du document de départ et peuvent être aussi longues que souhaitées, mais aussi être classées selon différents critères, grâce à des codes couleurs ou des rassemblements par fichiers. Parmi les logiciels permettant ce type de manipulations se trouvent Pundit, pour annoter des pages web directement depuis le navigateur, mais aussi l'Annotation Studio du MIT qui fonctionne à partir de textes en différents formats (Word, PDF, HTML...) ou encore CATMA (*Computer Assisted Textual Markup and Analysis*) de l'université d'Hambourg qui présente l'avantage d'être couplé avec un outil d'analyse et un autre de visualisation des résultats obtenus.
- 32 L'annotation peut être manuelle, mais également automatisée, notamment grâce à la technologie de reconnaissance d'entités nommées (Named Entity Recognition) qui à partir d'un référentiel va être en mesure de reconnaître automatiquement des éléments d'un corpus et de leur attribuer une étiquette adaptée (Tous les adjectifs, les noms de lieux etc.). GATE (*General Architecture for Text Engineering*) fait partie des logiciels qui possèdent ce type de fonctionnalités. Dans le cadre du projet intitulé *Political Futures Tracker* relatif aux élections législatives de 2015 en Grande-Bretagne, GATE a ainsi été utilisé pour identifier les thématiques abordées dans un corpus de tweets de candidats. Pour chaque thématique (Education, emploi, environnement...), un référentiel de termes a été élaboré manuellement puis le logiciel a permis d'automatiser le classement de milliers de tweets par sujet. Ceci révèle par exemple que dans la thématique liée à l'imposition, les candidats conservateurs ont davantage évoqué la taxe professionnelle et les possibilités de déduction fiscale tandis que les travaillistes se focalisaient sur des questions liées à l'évasion fiscale et à la TVA.
- 33 Outre les logiciels d'annotation généralistes précédemment évoqués, il existe également des outils plus spécialisés, comme le *UAM Corpus Tool*, recommandé pour des travaux linguistiques, ou encore Praat, utilisé en phonétique pour l'annotation de bandes sons. En effet, les outils de ce type peuvent être employés pour le traitement de sources non-textuelles. ANVIL ou Advane permettent par exemple d'annoter des documents vidéo tandis que Ratsnake est destiné au traitement d'images.

- 34 En complément de ces outils ou de manière autonome peuvent être utilisés des logiciels d'exploration de données qui vont aider à produire du sens à partir d'un corpus généralement conséquent en rendant plus visibles des motifs (*Patterns*) ou tendances difficilement perceptibles par des moyens traditionnels. Pour une introduction détaillée au principe d'exploration de données, consulter par exemple l'ouvrage *Data Mining, Concepts and Techniques* (Han, Kamber & Pei, 2012).
- 35 Ainsi, le logiciel TXM développé par une équipe de l'ENS de Lyon peut notamment procéder à de l'étiquetage morphosyntaxique mais aussi proposer des informations sur le corpus textuel étudié comme le classement de mots par ordre d'occurrence, des listes de cooccurrences, de collocations ou de spécificités statistiques. Pour aider à l'interprétation de ces résultats, des modules de visualisation sont souvent ajoutés à ces outils, ce qui est le cas pour CATMA, déjà évoqué, mais aussi pour TXM ou encore pour la suite Voyant créée par Stefan Sinclair et Geoffrey Rockwell. Celle-ci fournit en effet la fréquence de mots recherchés, leur contexte d'apparition dans le corpus mais aussi des graphiques, des lignes de bulles, des mandalas générés à partir du corpus. Pour des visualisations plus élaborées, des logiciels comme Tableau Desktop ou Orange de Biolab sont utilisés tandis que des alternatives spécialisées sont également disponibles, telles que Gephi, particulièrement performant en matière de visualisation de réseaux ou VisualEyes combinant visualisations cartographiques et chronologiques.

## Cycle de recherche et numérique : Collaboration et diffusion

- 36 Ces visualisations sont des supports d'analyse mais aussi de diffusion des résultats de ces analyses auprès de collaborateurs proches, de pairs, d'agences de financement ou du grand public. La collaboration et la diffusion sont les deux derniers volets de ce panorama des outils numériques d'aide à la recherche. En effet, des fonctionnalités facilitant un travail à plusieurs sur un support commun sont intégrées à un nombre croissant de logiciels mais certains sont spécifiquement conçus pour ce type d'approche, qu'il s'agisse de partage de documents, de gestion de projets collectifs, ou de *crowdsourcing*. Ce terme, initialement employé dans le domaine de l'entreprise et aujourd'hui appliqué bien au-delà est composé des mots *crowd* et *outsourcing* et désigne le fait de confier des tâches généralement réalisées par des professionnels en interne à un public extérieur, souvent amateur. Ce terme est aujourd'hui appliqué à des initiatives très diverses, de la participation à l'élaboration de contenus en ligne sur le modèle de Wikipedia, à la rédaction d'articles pour des blogs, à l'étiquetage d'images sur Flickr, ou à l'élaboration de logiciels open-source.
- 37 En matière de recherche, le *crowdsourcing* se manifeste le plus souvent sous la forme d'appels à participation lancés par des institutions patrimoniales ou des équipes pour la réalisation de tâches chronophages mais ne nécessitant pas de compétences complexes. Ainsi, pour son projet TROVE, la bibliothèque nationale d'Australie a fait appel à des bénévoles suite à un projet ambitieux de numérisation de certaines parties de ses collections afin de relire les documents scannés puis traités par un logiciel de reconnaissance de caractères pour les classer mais aussi corriger les coquilles, apporter commentaires et précisions ou encore télécharger des documents annexes pertinents. Pour faciliter ces pratiques s'inscrivant dans la lignée des sciences participatives, des plateformes dédiées ont vu le jour comme Scripto, Zooniverse ou Crowdcrafting. Cette

perspective collaborative est également présente dans les sites de témoignages évoqués précédemment ou du type de Storycorps qui depuis 2015, grâce à son application pour téléphone portable offre un support en ligne hébergeant des centaines de récits personnels provenant d'internautes essentiellement nord-américains. Elle se manifeste également en traduction sur la plateforme Zanata par exemple.

- 38 Cette utilisation croissante du *crowdsourcing* s'inscrit souvent en matière de recherche dans un mouvement plus large en faveur d'une science dite « ouverte ». Bien que la définition de l'expression ne soit pas consensuelle (Voir Fecher & Friesike, 2014), il apparaît cependant que l'évolution technologique puisse dans certains cas faciliter une interaction de la communauté scientifique avec des interlocuteurs plus nombreux et aux profils plus hétérogènes, qu'il s'agisse de collaboration comme évoquée plus haut mais aussi de diffusion de « livrables » plus variés qu'auparavant. En effet, il est désormais possible par exemple de partager des contenus alors que le processus de recherche n'a pas atteint son terme, tels que des bibliographies commentées sur WorldCat, la base de données bibliographiques de l'OCLC (*Online Computer Library Center*), ou encore des jeux de données via des dépôts (*Repositories*) comme celui du projet Dataverse ou de Figshare. De manière significative, ces dépôts délivrent des identifiants permettant de citer les contenus déposés au même titre que des publications classiques, laissant entrevoir l'éventualité à terme d'une reconnaissance officielle de ce type de résultats.
- 39 Des plateformes inspirées des carnets de recherche des anthropologues où chaque étape du travail scientifique est documentée ont également vu le jour, à l'image d'Hypothèses, portée par le centre pour l'édition électronique ouverte (CLEO). Résultats intermédiaires, réflexions méthodologiques, diapositives de colloques, demandes de conseils techniques etc. y trouvent un espace de diffusion, espace également propice à réduire l'effet « tiroir » (*File Drawer Effect*, Rosenthal, 1979), expression désignant la propension à ne publier qu'à propos d'études ayant abouti à un résultat estimé satisfaisant et consignait à l'oubli nombres de travaux dont le cheminement serait néanmoins précieux pour les chercheurs travaillant sur des problématiques identiques.
- 40 Enfin, des alternatives aux colloques traditionnels se multiplient à l'instar des THATCamps par exemple. Les créateurs de ce concept de manifestation scientifique né en 2008 se sont inspirés des *bootcamps* pratiqués en informatique et le qualifient de *unconference*, mettant l'accent sur son caractère délibérément ouvert à tous au-delà de la sphère scientifique, transdisciplinaire et privilégiant les partages d'expérience et de compétences informels. Un camp a ainsi été organisé à l'université Grenoble Alpes en 2018 sur la thématique de l'utilisation des données ouvertes et logiciels libres dans la recherche en SHS. Un THATCamp Shakespeare s'est tenu à l'université de Géorgie en 2017 tandis qu'un autre organisé à l'université de Caroline du Nord en 2016 s'intéressait au militantisme en ligne de mouvements tels qu'Occupy Wall Street ou Black Lives Matter.

## Projets numériques en études anglophones

- 41 Les outils numériques au service de la recherche ont donc connu une croissance exponentielle ces dernières années et gagné en fonctionnalités, ouvrant la voie à des projets cherchant à exploiter les opportunités offertes. Le projet *Data Mining with Criminal Intent* (Roy Rosenzweig Center for History and New Media, Bailey Online,

TAPoR, Digging into Data) illustre cette combinaison d'approches et d'outils mobilisés pour répondre à une problématique de recherche, ici déterminer le mode d'empoisonnement le plus courant au Royaume-Uni à partir d'un corpus d'environ 200.000 descriptifs de procès rédigés entre 1674 et 1913. Voici un résumé simplifié de la démarche adoptée : grâce à un moteur de recherche, les notices relatives aux cas d'empoisonnement ont d'abord été identifiées puis téléchargées via Zotero. Voyant ensuite été employé pour éliminer des notices les mots ne présentant pas d'intérêt pour l'analyse (*stop-words*, terminologie légale...) puis pour référencer par collocation les termes associés au mot « poison ». Il ressort de ce travail que les boissons sont le mode d'empoisonnement le plus fréquent sur l'ensemble du corpus étudié.

- 42 Ian Lancashire, de l'université de Toronto, a quant à lui conduit grâce à des outils de textométrie notamment, une étude longitudinale des évolutions lexicales et syntaxiques de l'œuvre de trois auteurs britanniques afin d'analyser les corrélations entre phénomènes langagiers (Appauvrissement du vocabulaire employé, répétitions, nombre de clauses par énoncés...) et apparition de la maladie d'Alzheimer (Le, 2011).
- 43 Les outils numériques peuvent ainsi être utilisés de façon combinée ou plus isolée. Dans le cadre du projet intitulé *Mapping the Republic of Letters* de l'université Stanford initié en 2007, des outils de visualisation cartographique ont été employés pour révéler les caractéristiques de la correspondance de célèbres auteurs du Siècle des Lumières tels que Voltaire ou pour la sphère anglophone, John Locke ou Benjamin Franklin. Les métadonnées de centaines de lettres envoyées et reçues par ces auteurs (date, destinataire, pays...) ont ainsi été saisies afin de produire des cartes illustrant leurs réseaux de correspondance pour tenter d'identifier des tendances qu'il aurait été difficile de percevoir à partir d'un référencement manuel notamment du fait de leur volume. Quels liens Franklin entretenait-il avec ses homologues européens, qui étaient ses interlocuteurs privilégiés, les pays avec lesquels il avait les contacts les plus étroits ? Quel parallèle peut être établi entre ces courriers et les événements de politique internationale de l'époque ?
- 44 Des chercheurs de l'université de Glasgow ont eux mis au point un outil (*Mapping Metaphor*) destiné à dresser un inventaire de métaphores employées dans un corpus de plus de 800.000 mots anglais sur une période d'environ 1300 ans tirés de l'*Historical Thesaurus of English*. Le traitement automatique appliqué au corpus permet ainsi de faire apparaître des liens entre des termes et des conceptions métaphoriques agencées par catégories, présentés sous forme de visualisation circulaire interactive. Ainsi, la notion d'intelligence apparaît comme souvent associée en anglais à des métaphores liées au goût autour de mots tels que *salty* ou *insipid* (Bagli, 2016) tandis que celles de la catégorie « Oiseaux » renvoient fréquemment à des émotions comme la fierté, le courage ou la peur (Muller, 2015).
- 45 Le projet DocuScope conduit par Michael Witmore de la *Shakespeare Folger Library* et Jonathan Hope de l'université Strathclyde permet lui d'automatiser la comparaison de différentes pièces de l'auteur d'un point de vue linguistique, rhétorique et structurel afin de faire apparaître les éléments de similitudes et de contrastes entre comédies et tragédies par exemple.

## Conclusion

- 46 Ces projets ambitieux et complexes nécessitent néanmoins l'agrégation de compétences diverses qui renvoient au dialogue interdisciplinaire évoqué par Dacos et Mounier en introduction. Le projet PHEME coordonné par l'université de Sheffield qui vise à construire un outil numérique capable de déterminer la véracité d'une information véhiculée sur les réseaux sociaux repose ainsi sur la collaboration de spécialistes du traitement automatique des langues, de l'exploration de données, de l'analyse de réseaux et de linguistes. Ceci ne signifie cependant pas nécessairement que l'utilisation d'outils numériques ne soit pas envisageable et porteuse scientifiquement à une échelle plus réduite, voire individuelle.
- 47 Se pose néanmoins la question de l'accessibilité technique de tels outils pour des chercheurs issus de cursus dépourvus d'enseignements dédiés. En 1968, l'historien Emmanuel Le Roy Ladurie sous-titrait un article rédigé pour le *Nouvel Observateur* de la phrase suivante : « l'historien de demain sera programmeur ou ne sera pas ». L'impact de cette affirmation est encore perceptible puisqu'un site lancé en 2012 proposant des tutoriels à destination de la communauté de recherche en SHS dans les sphères francophones, anglophones et hispanophones porte le nom de *Programming Historian*. Pour autant, ce profil hybride reste relativement rare, notamment en France, non seulement en histoire mais dans l'ensemble des SHS. Paul Bertrand soulignait ainsi lors du THATCamp qui s'est tenu à Paris en 2012 la difficulté de trouver temps et énergie pour acquérir et entretenir ce type de compétences dans une « vie d'historien, déjà bien remplie en soi, une vie de recherche ». Toutefois, comme le rappelle Emilien Ruiz, co-auteur du site 'La boîte à outils des historien(ne)s', depuis la célèbre citation de Ladurie, des interfaces ont été créées pour se placer entre l'homme et la machine et ainsi réduire drastiquement le besoin de compétences en programmation de la part des chercheurs (Ruiz, 2011). L'ensemble des outils évoqués plus haut est ainsi accessible sans aucune compétence de ce type tout en offrant des fonctionnalités approfondies. Pour les prendre en main et éviter leurs pièges potentiels, une phrase de formation ou d'autoformation reste indispensable. Il n'en demeure pas moins que les prérequis techniques pour leur utilisation ont indéniablement baissé au cours des dernières décennies, permettant aux chercheurs en SHS d'acquérir pour la réalisation de nombreuses tâches une autonomie croissante et ouvrant la voie à des pistes scientifiques prometteuses.
- 48 Pour autant, cela ne signifie pas que les méthodes et outils antérieurs soient nécessairement rendus obsolètes par ces innovations, comme le suggèrent les travaux de chercheurs tels que Franco Moretti, célèbre et polémique historien de la littérature à l'origine du concept de *distant reading* pour qui l'analyse informatiquement automatisée d'un volume de textes allant au-delà des capacités humaines de traitement doit remplacer la pratique de lecture rapprochée généralement utilisée par les spécialistes de la discipline. En effet, si le débat en la matière reste vif et si toute conclusion ne peut être que temporaire, le consensus qui se dessine va davantage dans le sens d'une complémentarité des approches (Cabral, Amado Laurel & Schuerewegen, 2014), ce que Moretti lui-même laisse entrevoir dans son étude consacrée à Hamlet et intitulée *Network Theory, Plot Analysis* (Moretti, 2011).
- 49 Le positionnement individuel ou institutionnel du chercheur sur ces questions est variable, allant d'une adhésion sans faille à un déterminisme technologique menaçant

d'obsolescence quiconque n'emprunterait pas cette voie à une volonté parfois farouche de défendre des pratiques antérieures ayant souvent fait leurs preuves contre toute rhétorique associant systématiquement et sans nuance évolution technologique et progrès. Il semble toutefois indéniable que l'irruption du numérique dans la recherche, en études anglophones comme dans la plupart des domaines scientifiques, ne peut plus être cantonnée à un effet de mode temporaire à l'initiative de quelques technophiles évoluant à la marge. Sans nécessairement adopter à titre personnel ces outils ou approches qui n'ont d'ailleurs pas vocation à répondre à toute problématique pour tout public en toute situation, ajouter les productions issues de la sphère des humanités numériques à la veille scientifique conduite par chacun dans sa discipline de spécialité ne peut cependant qu'enrichir la réflexion de tous. Pour citer la célèbre phrase d'introduction à l'ouvrage « *A New Guide to Science* » publié par Isaac Asimov en 1984, « *Almost in the beginning was curiosity* ».

---

## BIBLIOGRAPHIE

- Asimov, Isaac. *Asimov's New Guide to Science*. Basic Books, 1984.
- Bagli, Marco. « Salt, Taste and the Conceptualization of Intelligence ». *Mapping Metaphor*, 24 mars 2016, <https://mappingmetaphor.arts.gla.ac.uk/blog/?p=358>.
- Barats, Christine (dir.), « Manuel d'analyse du web en sciences humaines et sociales ». *Lectures*, Décembre 2013. <http://journals.openedition.org/lectures/13030>.
- Berra, Aurélien. « Faire des humanités numériques ». *OpenEdition Press*, 2012, p. 25-43.
- Bauer Sylvie, Larssonneur Claire, Machinal, Hélène et Regnauld Arnaud, *Subjectivités numériques et posthumain*. Presses Universitaires de Rennes. À paraître (2019)
- Bertrand, Paul. « Les digital humanities sont-elles solubles dans le Steampunk ? : Conférence inaugurale ». *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*, Éditions de la Maison des sciences de l'homme, 2012. <http://books.openedition.org/editionsmsmh/327>.
- Bourdaloie, Hélène. « Ce que le numérique fait aux sciences humaines et sociales. épistémologie, méthodes et outils en questions ». *Tic&société*, n° Vol. 7, N° 2, mai 2014. [journals.openedition.org](http://journals.openedition.org/ticetsociete.1500), doi:10.4000/ticetsociete.1500.
- Boyd, Juliet. « Creating Database Forms ». 12 juillet 2014, <https://julietboyd.wordpress.com/2014/07/12/creating-database-forms/>.
- Burnett, Cathy, et Guy Merchant. « Is There a Space for Critical Literacy in the Context of Social Media? » *English Teaching: Practice and Critique*, vol. 10, n° 1, mai 2011, p. 41-57.
- Clavert, Frédéric, et al. *Expérimenter les Humanités numériques*. Presses Universitaires de Montréal, 2017.
- Crompton, Constance, et al. *Doing Digital Humanities: Practice, Training, Research*. Taylor & Francis, 2016.

- Dacos, Marin, et Pierre Mounier. *Humanités numériques*. Institut français, mars 2015, p. 9782354761097. <https://hal.archives-ouvertes.fr/hal-01228945>.
- Fecher, Benedikt, et Sascha Friesike. « Open Science: One Term, Five Schools of Thought ». *Opening Science*, Springer, Cham, 2014, p. 17-47.
- Han, Kamber et Pei. *Data mining: concepts and techniques*. Elsevier, 2012.
- Gardiner, Eileen, et Ronald G. Musto. *The Digital Humanities: A primer for Students and Scholars*. Cambridge University Press, 2015.
- Greenspan, Brian. « Are Digital Humanists Utopian ? » *Debates in the Digital Humanities*. 2016. <http://dhdebates.gc.cuny.edu/debates/text/84>
- Greetham, David. « The Resistance to Digital Humanities ». *Debates in the Digital Humanities*, 2012. <http://dhdebates.gc.cuny.edu/debates/text/47>
- Guichard, Éric. « L'internet et les épistémologies des sciences humaines et sociales ». *Revue Sciences/Lettres*, n° 2, 2014.
- Jänicke, Stefan, et al. « On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges ». *Eurographics Conference on Visualization (EuroVis)-STARs*. The Eurographics Association, 2015.
- Lacaze, Grégoire, et Aix-Marseille Université. *Du discours rapporté dans les tweets : reconfiguration des pratiques citationnelles du discours circulant*. Colloque Ci-dit. Université Libre de Bruxelles. 2018.
- Larsonneur, Claire (dir) « Digital Subjectivities ». *Angles*, vol. 7, juin 2018, <http://angles.saesfrance.org/index.php?id=954>.
- Larsonneur, Claire, Regnaud, Arnaud, Cassou-Noguès, Pierre, Touiza, Sara (Eds). *Le Sujet Digital*. Les presses du réel. 2015
- Laurel, M. H. A., et al. *Lire de près, de loin : Close vs Distant Reading*. Classiques Garnier, 2014.
- Le Roy Ladurie, Emmanuel. *La fin des érudits : l'historien de demain sera programmeur ou ne sera pas*. Le nouvel observateur, 8 mai 1968.
- Le, Xuan, et al. « Longitudinal Detection of Dementia through Lexical and Syntactic Changes in Writing: A Case Study of Three British Novelists ». *Literary and Linguistic Computing*, vol. 26, n° 4, décembre 2011, p. 435-61.
- Lemny, Stefan. « « L'historien de demain sera programmeur... » : Emmanuel Le Roy Ladurie et les défis de la science ». *L'Histoire à la BnF*, 28 décembre 2017. <https://histoirebnf.hypotheses.org/1505>.
- MacKenzie. *Writing about Reading about Writing about Reading*. <http://operation-critique.tumblr.com/>.
- Mair Jones, Ffion, « A New Look at the Correspondence of Thomas Pennant and Richard Bull (1783-1798) ». *Curious Travellers Editions' Website*. [https://editions.curioustravellers.ac.uk/pages/new\\_look\\_Pennant\\_Bull.html](https://editions.curioustravellers.ac.uk/pages/new_look_Pennant_Bull.html)
- Moretti, Franco. *Distant reading*. Verso Books, 2013.
- Moretti, Franco. « Network theory, plot analysis ». *New Left Review*, 2011.
- Muller, Sarah. « Birds, Pride, Fear and Courage ». *Mapping Metaphor*, 3 juillet 2015. <https://mappingmetaphor.arts.gla.ac.uk/blog/?p=329>.



Notay, Balviar. « CORE Becomes the World's Largest Aggregator ». *Jisc Scholarly Communications*, 1<sup>er</sup> juin 2018. <https://scholarlycommunications.jiscinvolve.org/wp/2018/06/01/core-becomes-the-worlds-largest-aggregator/>.

Nyhan, Julianne, et al. *Defining Digital Humanities: A Reader*. Ashgate Publishing Limited, 2013.

Presner, Todd, Schnapp Jeffrey et al. *Digital Humanities Manifesto 2.0*. Humanitiesblast. 2009. [http://www.humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf)

Rosenthal, Robert. « The file Drawer Problem and Tolerance for Null Results. » *Psychological Bulletin*, vol. 86, n° 3, 1979, p. 638.

Ruiz, Émilien. « Les historiens seront-ils finalement programmeurs ? » *La boîte à outils des historien-ne-s*, 22 septembre 2011. <http://www.boiteaoutils.info/2011/09/les-historiens-seront-ils-finalement/>.

Schulz, Kathryn. « The Mechanic Muse - What Is Distant Reading? » *The New York Times*, 24 juin 2011. <https://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html>.

St. Onge, Ruth-Ellen. « Compte rendu de Moretti (Franco), *Distant Reading*. London/New York, Verso, 2013, 244 p. » *CONTEXTES. Revue de sociologie de la littérature*, mars 2014. <http://journals.openedition.org/contextes/5870>.

Terras, Melissa. « Peering Inside the Big Tent Digital Humanities and the Crisis of Inclusion. Keynote. » Presented at: *Interface 2011, UCL*. 2011. [discovery.ucl.ac.uk](http://discovery.ucl.ac.uk/1355076/), <http://discovery.ucl.ac.uk/1355076/>.

Welger-Barboza, Corinne. « Les digital humanities aujourd'hui : centres, réseaux, pratiques et enjeux ». *Read/Write Book 2 : Une introduction aux humanités numériques*, Pierre Mounier Ed, OpenEdition Press, 2012. <http://books.openedition.org/oep/244>.

### Archives, outils et projets référencés

« Accents and Dialects. British Library - Sounds ». <https://sounds.bl.uk/Accents-and-dialects>.

« Advene for Video Annotation ». <https://www.advene.org/>.

« Amérique/Europe : Les humanités numériques en partage ? Enjeux, innovation et perspectives. 15e colloque annuel de l'institut des Amériques », 2017. [http://www.institutdesameriques.fr/sites/default/files/documents\\_actualites/colloqueidaprogramme\\_bd.pdf](http://www.institutdesameriques.fr/sites/default/files/documents_actualites/colloqueidaprogramme_bd.pdf).

« Annotation Studio ». <https://www.annotationstudio.org/>.

« ANVIL: The Video Annotation Research Tool ». <http://www.anvil-software.org/>.

« BASE - Bielefeld Academic Search Engine ». <https://www.base-search.net/about/fr/index.php?>

« Between Two Worlds: Poetry and Translation. British Library - Sounds ». <https://sounds.bl.uk/Arts-literature-and-performance/Between-two-worlds-poetry-and-translation>.

« CATMA – For Undogmatic Textual Markup and Analysis ». <http://catma.de/>.

« CEISMIC (Canterbury Earthquake Digital Archive) ». <http://www.ceismic.org.nz/>.

« CLARIN - European Research Infrastructure for Language Resources and Technology ». <https://www.clarin.eu/>.

« CLiC Dickens ». <https://www.birmingham.ac.uk/schools/edacs/departments/englishlanguage/research/projects/clic/index.aspx>.

- « CoRD | Zurich English Newspaper Corpus (ZEN) ». <http://www.helsinki.fi/varieng/CoRD/corpora/ZEN/>.
- « CORE ». <https://core.ac.uk/>.
- « Corpus of Welsh Language Tweets | ». <http://techiaith.cymru/corpora/twitter/?lang=en>.
- « Data Mining with Criminal Intent: Using Zotero and TAPoR on the Old Bailey Proceedings ». *DHI* (blog), 8 avril 2010. <https://www.dhi.ac.uk/projects/criminal-intent/>.
- « Ethique et droit ». *Hypothèses* (blog). <https://ethiquedroit.hypotheses.org/>.
- « Figshare ». <https://figshare.com/>.
- « FreeDict ». <https://freedict.org/>.
- « GATE (General Architecture for Text Engineering) ». <https://gate.ac.uk/>.
- « Gephi - The Open Graph Viz Platform ». <https://gephi.org/>.
- « Human Rights' Web Archive ». Center for Human Rights' Documentation and Research at Columbia University, s. d. <https://hrwa.cul.columbia.edu/>.
- « Internet Archive: Wayback Machine ». <https://archive.org/web/>.
- « Mapping Metaphor: Home ». <https://mappingmetaphor.arts.gla.ac.uk/>.
- « Mapping the Republic of Letters ». <http://republicofletters.stanford.edu/>.
- « Mendeley - Reference Management Software & Researcher Network ». <https://www.mendeley.com/homepage6/?switchedFrom=>.
- « Newcastle Electronic Corpus of Tyneside English ». <https://research.ncl.ac.uk/necte/>.
- « NZETC (New Zealand Electronic Text Collection) ». <http://nzetc.victoria.ac.nz/tm/scholarly/tei-NZETC-About-accessingtexts.html>.
- « Observing the 1980s - Oral history | British Library - Sounds ». <https://sounds.bl.uk/Oral-history/Observing-the-1980s>.
- « Open Refine ». <http://openrefine.org/>.
- « Orange for data mining ». <https://orange.biolab.si/>.
- « oXygen Visual TEI Editor ». [https://www.oxygenxml.com/xml\\_editor/tei\\_editor.html](https://www.oxygenxml.com/xml_editor/tei_editor.html).
- « PHEME ». <https://www.pHEME.eu/>.
- « Praat: doing Phonetics by Computer ». <http://www.fon.hum.uva.nl/praat/>.
- « Project Lost & Found ». *DELAMAN (Digital Endangered Languages and Musics Archive Network)* (blog). <http://www.delaman.org/project-lost-found/>.
- « Pundit - Semantic Web Annotation ». Pundit. <http://thepund.it/semantic-web-annotation/>.
- « Ratsnake Annotation Tool ». <https://is-innovation.eu/ratsnake/>.
- « The Berkeley Revolution ». <http://revolution.berkeley.edu/about/>

## NOTES

1. Le sujet digital, Labex Arts-H2H, co-piloté par Claire Larssonneur, Arnaud Regnauld, Pierre Cassou-Noguès, 2012-16, <http://sujetdigital.labex-arts-h2h.fr/fr>

2. Post-humain et subjectivités numériques, Centre culturel international de Cerisy, 23-30 juin 2016, <http://www.ccic-cerisy.asso.fr/posthumain16.html>
  3. Cf « Wayback Machine General Information », *Internet Archive Help Center*, Internet Archive. Updated August 2018. Visited 22 February, 2019.  
<<https://help.archive.org/hc/en-us/articles/360004716091-Wayback-Machine-General-Information>>
  4. Les logiciels évoqués ici le sont à titre d'exemple des fonctionnalités décrites mais leur vitesse de renouvellement étant élevée, d'autres les auront peut-être remplacés d'ici publication. Sauf indication contraire, il s'agit de logiciels disponibles gratuitement.
- 

## RÉSUMÉS

Bien que l'utilisation de technologies alors perçues comme novatrices dans la recherche en sciences humaines et sociales ne soit pas un phénomène particulièrement récent, le champ des humanités numériques tel que défini en France par Dacos et Mounier (2014) ou Bourdeloie (2014) est encore en construction. Potentiel, origines, contours, méthodes continuent de faire l'objet de nombreux débats.

En parallèle cependant, des projets voient le jour dans un nombre croissant de pays et de disciplines. Pour autant, malgré quelques initiatives notables, les études anglophones françaises restent à la marge de ce mouvement. Les raisons sont multiples, mais l'une d'entre elles réside indéniablement dans une difficulté pour les chercheurs du domaine à percevoir l'apport éventuel de l'introduction du numérique dans des pratiques existantes reposant sur des approches ayant démontré leur valeur scientifique.

Cet article se propose donc d'offrir un panorama synthétique des contributions possibles des ressources et outils numériques aux différentes étapes du travail de recherche : état de l'art, préparation et constitution de corpus, manipulation de données, analyse, travail collaboratif et diffusion. Cette introduction sera illustrée d'exemples tirés des disciplines des études anglophones françaises afin d'informer la réflexion des chercheurs du domaine sur l'existant en la matière et d'offrir une première réponse à leurs possibles interrogations sur la pertinence d'intégrer les pistes évoquées ici à leur propre démarche de recherche.

Although the use of technologies previously seen as innovative in humanities and social sciences research is not a particularly recent phenomenon, the field of the digital humanities as defined in France by Dacos and Mounier (2014) or Bourdeloie (2014) is still under construction. Its origins, outline, methods and potential continue to be the subject of much debate.

However, alongside those ongoing debates, projects are emerging in an increasing number of countries and disciplines. Nevertheless, despite some promising initiatives, English studies in France remain at the margins of this movement. There are many reasons but one of them is undeniably due to the difficulty for researchers in the field to perceive the possible contribution of digital technology to existing practices generally based on approaches which have demonstrated their scientific value.

This article therefore aims at offering a synthetic overview of the potential input of digital resources and tools to the various stages of the research cycle: state of the art, corpus preparation and constitution, data manipulation, analysis, collaborative work and dissemination. This introduction will be illustrated by examples drawn from the disciplines of English studies in

France in order to help researchers in the field approach the sometimes complex and intimidating area of the digital humanities and try to answer their questions on the relevance of integrating some of the elements mentioned here into their own research.

## INDEX

**Keywords :** Digital Humanities, English Studies, Tools, Resources, Projects, Corpus, Data, Mining, Software, Websites

**Mots-clés :** humanités numériques, études anglophones, outils, ressources, projets, corpus, données, exploration, logiciels, sites

## AUTEUR

**GÉRALDINE CASTEL**

Maître de conférences

Université Grenoble Alpes

geraldine.castel@univ-grenoble-alpes.fr